

Visual validation of the appearance of chromatic objects rendered from spectrophotometric measurements

KHALIL HURAI BAT,^{1,*} ESTHER PERALES,¹ ERIC KIRCHNER,² IVO VAN DER LANS,²
ALEJANDRO FERRERO,³ AND JOAQUÍN CAMPOS³

¹Department of Optics, Universidad de Alicante, Alicante, Spain

²AkzoNobel Technology Group Color, Sassenheim, The Netherlands

³Instituto de Óptica, Consejo Superior de Investigaciones Científicas, Madrid, Spain

*Corresponding author: khalil.huraibat@ua.es

Received 3 November 2020; revised 4 January 2021; accepted 8 January 2021; posted 8 January 2021 (Doc. ID 413890); published 5 February 2021

We validate a physically based and spectral rendering framework with improved color reproduction. With a recently developed model, we take into account both the colorimetric specifications of the rendering display as well as the spectral and angular characteristics of lighting and also the spectral reflectance of the objects. Therefore, it should provide much better color reproduction than those based on the common standard red, green, blue (sRGB) color space. In addition, it allows real-time rendering on modest hardware and displays. We evaluated the color reproduction of the new rendering framework by psychophysical tests using spectrophotometric measurements of 30 chromatic paint samples. They were rendered on an iPad display, as viewed inside the Byko-spectra effect light booth. We asked 16 observers to evaluate the color match by directly comparing the rendered samples with the physical samples, using two different psychophysical assessment methods. The color reproduction was found to be strongly improved with respect to results obtained with default sRGB color encoding space. The average color reproduction match was found to be equivalent to $\Delta E_{00} = 1.6$, which is a small but noticeable color difference. In 80% of the visual assessments, the color reproduction was described as being at least as good as between “difference visible but still acceptable” and “difference visible, doubtful match.” © 2021 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

<https://doi.org/10.1364/JOSAA.413890>

1. INTRODUCTION

The rendering of different types of materials has developed greatly over the past decades [1–4]. Current commercial renderers provide frames (images) that serve the needs for applications such as the cinema and games industries [5]. However, when compared to real-world objects, the rendered images are not accurate enough for more critical applications such as automotive design, especially for complex coatings such as, for instance, iridescent with effect pigments [6,7]. The reason is that the dependence of the color on viewing and illumination directions is not covered correctly in current rendering algorithms [8]. In addition, complex textures such as sparkle or graininess are still difficult to render in a convincing way [9,10]. Both phenomena are either absent or not represented well in current rendering software [10–13]. However, rendering software has matured over the past decades. Currently, several rendering software packages are commercially and open-source available, such as Maya, Keyshot, VRED (Autodesk), Radiance, Revit, and Mentalray (Nvidia) [5,14]. The rendered images created by

these software packages suggest (and sometimes even claim) a photorealistic quality, but surprisingly few articles analyze the visual match of these images with their corresponding real objects. Some of these few analyses show that in many cases the renderings are not realistic in terms of color match [11,15]. This mismatch gets worse for objects with complex reflectance, with both high spectral and angular dependence, as is with the case of iridescent effect coatings.

The first step to improve color match in rendering is to include a full spectral approach accounting for the spectral reflectance of the object and as the spectral power distribution (SPD) of the light sources. Several spectral renderers are currently available, such as Mitsuba [16], ImpastoR [17], ART, and Mental Ray. All of these renderers require specialist hardware, such as fast graphics cards, and would not render real-time on more modest hardware such as a tablet computer. Similar to earlier work, we will describe the lighting environment by global illumination models [15,18–21].

In addition, we need to improve the absolute color representation on displays. It is well-known that the device-independent

standard red, green, blue (sRGB) method, which is the default technique to calculate digital color representations, is often not accurate. The model parameters in the sRGB method were determined almost 30 years ago, when most displays were cathode-ray tube (CRT). Current displays are mostly based on organic LED (OLED) or LCD technology instead, which makes the use of the sRGB default model parameters for gamma and color primaries doubtful [22,23]. For accurate absolute color rendering, it is important to take into account the technical specifications of the display. We recently proposed the mobile display characterization and illumination model (MDCIM) [22]. It takes into account not only the technical specifications of LCD and OLED displays, but also the influence from ambient lighting.

This paper describes part of a research collaboration, in which we developed a physically based rendering framework for improving color and texture reproduction of car paints. We validated the performance of the framework using psychophysical methods, in which rendered images are visually compared with their corresponding real-world objects.

In the current paper, we provide the groundwork by limiting ourselves to uniform and chromatic paint samples. In coming work, we will extend the analysis to render color flop, graininess, and sparkle, where not only spectral but also spatial features can be included by spectrophotometric measurements and physics-based analytical models.

2. RENDERING METHODOLOGY

A. Rendering Pipeline

In order to avoid solutions that only run well on high-end personal computer (PC) or graphics cards, we developed the rendering pipeline on modest hardware, as exemplified by an iPad tablet computer, and based it on OpenGL 3.0 embedded systems (ES), an open-source graphics library that already has the basic functionality of rendering. OpenGL is also supported on Android devices and Windows PCs. We expect that our methodology is portable to other technology platforms [24]. We used edition 5 of the iPad (here abbreviated as iPad 5), which was commercially released in March 2017.

We turned the red, green, blue alpha (RGBA)-based framework into a fully spectral rendering pipeline by using 16 spectral bands in the visible range from 400 to 700 nm, with a spectral bandwidth of 20 nm. We process the spectral data repeatedly in blocks of four spectral bands through the RGBA pipeline of OpenGL 3.0 ES. Only at the final stage of the calculations did we combine all of the calculated spectral data into one resulting red, green, blue (RGB) image using the MDCIM model [22], so the final images are still in the conventional RGB format. The MDCIM parameter values used were already published in one of our earlier articles [22]. This approach has some similarity to Darling's [25], where they used a six-channel workflow in order to process calculations in real-time, which was found to result in accuracy issues for multispectral illumination input. In our approach, we use 16 bands, which should further improve color accuracy, and we took measures so we can still work in real time.

Rendered images cannot be more accurate in color than the color space used for color encoding. In common rendering



Fig. 1. Byko-spectra effect light booth (BYK-Gardner). The left image shows the inner structure of the light booth with the rotating platform.

software, the calculation of RGB images utilizes the device-independent sRGB method. Therefore, they do not take into account the technical specifications of the display on which the rendered images are shown. This introduces a substantial variation in displayed colors [23]. We recently developed a method to account for the display characteristics as an alternative to the sRGB method, the MDCIM method [22]. Common display calibration methods, such as Spyder and i1Pro, only make colors on displays consistent with sRGB color space but they do not make colors accurately represent surface colors under a variety of lighting conditions, which is why we use MDCIM, which considers the spectral irradiance of ambient light. Finally, we implemented spatial dithering to achieve further improvement in color reproduction.

This rendering pipeline needs spectral distributions for describing the lighting and the spectral reflectance of objects. The required spectral distributions are obtained by spectrophotometric measurements.

B. Virtual Light Booth

The perceived color of objects critically depends on the ambient light surrounding the objects, and it should be possible to integrate any of the usual lighting environments into this rendering framework. In order to use normalized lighting conditions, we selected those of the Byko-spectra effect light booth from supplier BYK-Gardner (see Fig. 1), a commercially available light booth that ensures well-defined, consistent, uniform, and repeatable lighting. This light booth is particularly suitable for visual inspection of effect coatings, allowing six standard viewing and illumination angles, which, according to international standards, are optimal for observing angular color variation of these coatings [26]. In addition, it includes adequate light sources to enable visual assessments of sparkle and graininess [26,27]. Both color variation and texture are targeted by this research in its broadest scope.

In order to render samples as they will be shown inside this light booth, we created the corresponding lighting environment according to the following multi-step approach that we developed over the past few years [24,27–29]:

Step 1: Geometrical model of the light booth

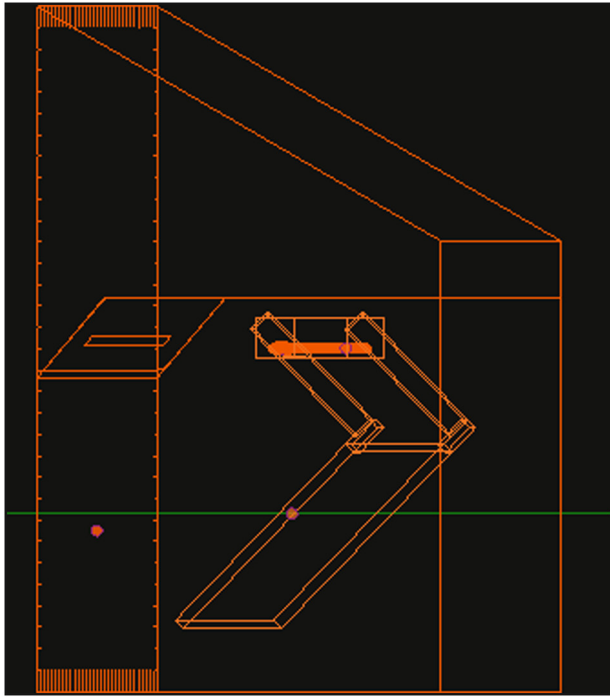


Fig. 2. Geometrical representation of the Byko-spectra effect light booth with its inner components as a 3D wireframe mesh in Blender.

We built a three-dimensional (3D) geometrical model of the light booth by using Blender, an open-source 3D modeling software widely used for creating 3D digital art and animations that allows a complete 3D creation pipeline to be designed [30]. The geometrical model consists of a 3D mesh containing a representation of each component of the light booth and their dimensions, as illustrated by the 3D model wireframe shown in Fig. 2, showing the box that forms the outer body of the light booth, the rotation platform where samples are placed, the light-source cavity, and the light-source tube itself. Since the light-source cavity is attached to the platform, the angle of illumination is fixed (to 45°) regardless of the rotation state. In the physical light booth, a lever enables the user to rotate the platform to select one of the six available illumination-viewing geometries (45as-15, 45as15, 45as25, 45as45, 45as75, and 45as110, where, according to the Deutsche Industrie Norm/American Society for Testing and Materials (DIN/ASTM) nomenclature, the first number specifies the illumination angle with respect to the sample normal, and the second one is with respect to the aspecular angle) [31,32]. We created in Blender six separate 3D geometrical models of the Byko-spectra light booth, corresponding to each of these six available geometries. In addition, we included both a rectangular aperture at the light-source cavity and a viewing slit of the light booth in the geometrical model. The rectangular aperture in the light-source cavity limits the width of the angular distribution of the luminous flux irradiating the objects, whereas the viewing slit limits that of the luminous flux reaching the observers.

Step 2: Spectral power distribution of the incident luminous flux

The light source inside the Byko-spectra effect light booth is a fluorescent light tube (Philips Master PL L90 De Luxe

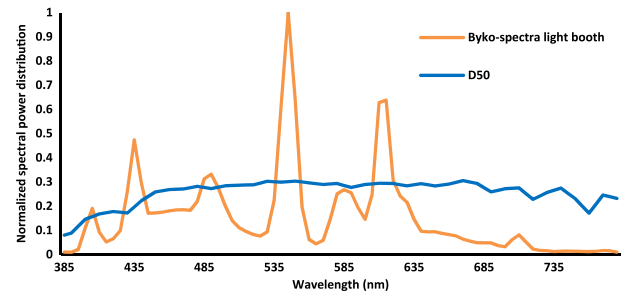


Fig. 3. Normalized spectral power distribution (SPD) of the Byko-spectra light booth together with the D50 illuminant.

55W/950/4p), with a SPD in principle similar to “daylight lighting.” We measured its spectral radiance with the tele-spectroradiometer Konica Minolta CS2000, using a Spectralon white sample, placed at exactly the same conditions as the test samples, as a reflectance transfer standard. Figure 3 shows the measured relative SPD together with the SPD of the standard D50 illuminant. It is observed that the measured spectrum is more peaked than the D50 illuminant, confirming previous work by Martínez-Verdú *et al.* [26]. They had concluded that the SPD of this fluorescent tube is not a good D65 simulator, but that it could be used to simulate D50. However, assuming that the same distribution might introduce color mismatch in some samples, we used the measured SPD in the rendering pipeline.

Step 3: Spatial and angular distribution of incident luminous source

The geometrical distribution of the incident luminous source impacts the illuminance on the samples and on the reflected luminous flux in the observer direction. The luminous intensity angular distributions of many luminaires are available in the IES/EULUMDAT format from the manufacturer. However, this is not the case for the light source inside the Byko-spectra light booth. In addition, since the distance between the light source and the sample is comparable to the length of the fluorescent lamp, we are not in far-field conditions, and the luminous intensity distribution can be used only as a rough approximation. Supported by spectrophotometric measurements, we modeled the spatial and angular distribution of luminous flux at the source and on the sample plane as follows.

We represent the two separate large fluorescent tubes by $N (=100)$ different point light sources along the long axis of each tube, and we assume that all point light sources have the same radiance intensity value (I_λ) and angular distribution. This assumption allows the value and distribution to be calculated from irradiance measurements on the sample plane. We measured this irradiance at $M (=55)$ different positions (E_λ) arranged in a grid, as illustrated in Fig. 4. The measured value of the irradiance (E_λ) at any position (x, y) is the resulting sum of irradiance contributions from each of the N point sources that constitute the light tubes. This is mathematically represented by Eq. (1),

$$E_\lambda(x, y) = \sum_{i=1}^N E_{\lambda,i} = \sum_{i=1}^N \frac{I_\lambda(\theta_i)}{d_i^2} \cos \theta_i, \quad (1)$$

where d is the distance between the light point source and the measurement point, and θ_i is the inclination angle with

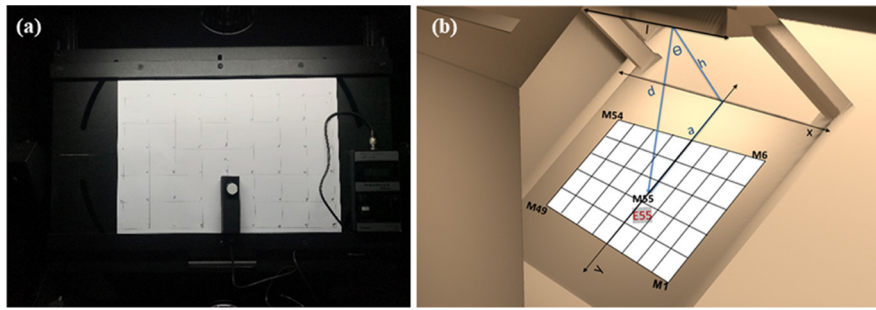


Fig. 4. Setup to computing the luminous intensity values (I_λ) from the measured illuminances (E_λ) at the intersection points (M) of the grid placed on the rotating platform, where di and θ_i are computed as $di = \sqrt{(x_i^2 + y_i^2) + b^2}$; $\theta_i = \tan \sqrt{x_i^2 + y_i^2} / b$.

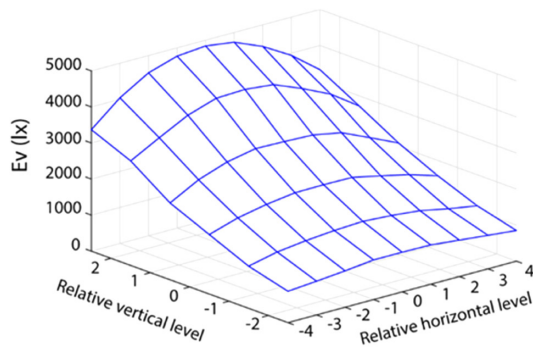


Fig. 5. Illuminance values measured on a grid of $M = 55$ points on the rotation platform of the Byko-spectra effect light booth.

respect to normal to the sample plane [Fig. 4(b)]. A function as in Eq. (2) is used to model the luminous intensity distribution of the point light sources, whose parameters σ and $I_{\lambda,0}$ are obtained by fitting to the measurements:

$$I_\lambda(\theta) = I_{\lambda,0} e^{-\left(\frac{\theta}{2\sigma}\right)^2}. \quad (2)$$

The measurements show that the illuminance varies considerably across the rotation platform (see Fig. 5). These results show that the illuminance is not uniform across the sample platform. As is already concluded by Martínez-Verdú *et al.*, the illuminance profile shows that this light booth is suitable for well-standardized visual observations only if relatively small samples are used [27].

The obtained value for σ is very large, meaning that the point light sources are best described by isotropic emission profiles. A comparison of the fitted model and the measured illuminance values shows that the predictions have a deviation of 10% on average. Since this deviation is much smaller than the measured inhomogeneity in illumination across the platform, we conclude that this model provides an important improvement for the purpose of this investigation. These results were integrated in the rendering framework by using an algorithm developed by Heitz *et al.* [33], which models polygonal light sources. Finally, we created an Illuminating Engineering Society (IES)/EULUMDAT file to describe the light environment for the rendering framework.

Step 4: Spectral reflectance of the light booth components

Through internal reflections, the inner surfaces of the light booth surrounding the test object might influence the final appearance of the test object, and they are integrated in the rendering pipeline. For inner walls and floor areas, our measurements show neutral colors and reflectance values of approximately 5%, confirming that the manufacturer of the light booth has minimized indirect light reflections, as expected from a standardized light booth.

Figure 6 shows an example of the final rendering of the Byko-spectra effect light booth, where we have placed a black hemisphere on the sample platform. Figure 6(a) is a view of the global illumination rendering, where the platform is shown from a side and ignores the viewing slit. In Fig. 6(b), we can see the specular reflection of the fluorescent tube in the surface of the hemisphere, as if seen through the viewing slit.

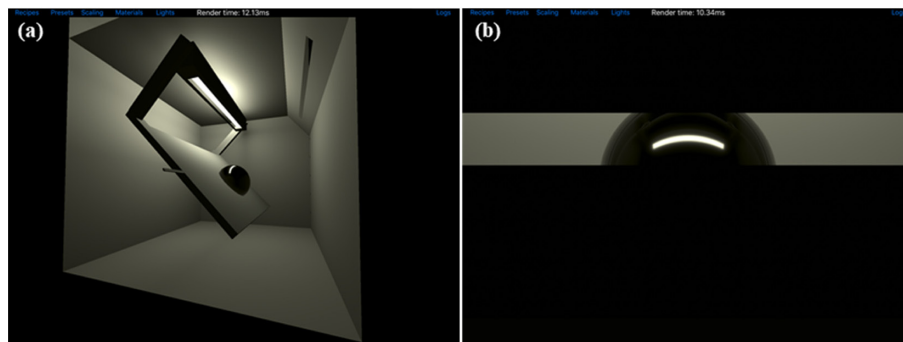


Fig. 6. Screenshots of the 3D rendering of the Byko-spectra effect light booth as shown on a tablet computer. (a) Side view of the light booth enhanced for this illustration. (b) Observer view through the viewing slit, actual rendering.

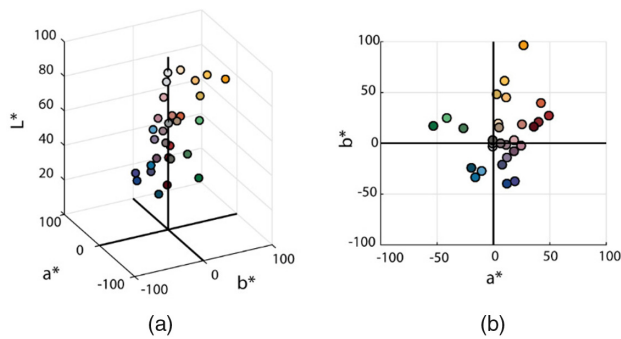


Fig. 7. Spectral reflectance data corresponding to the as 45 measurement geometry of the BYK mac-i for the samples studied in this article, represented (a) in CIE — $L^*a^*b^*$ color space, and (b) in the chromaticity a^* , b^* diagram.

3. VISUAL EXPERIMENT TO EVALUATE COLOR REPRODUCTION

We carried out visual tests to evaluate the color reproduction of the presented rendering framework. In these tests, observers are asked to directly compare the color of samples as viewed inside a real-world Byko-spectra effect light booth with the corresponding sample on an iPad display, rendered inside the virtual light booth. We use standard psychophysical methods to obtain quantitative results about the perceived color match. The visual experiment was designed to allow future investigations on color and texture of effect coatings.

We selected 30 highly glossy chromatic paint samples. In order to make the visual assessment of the quality of the color match as simple as possible for the observers, we used flat samples for this test. The samples cover a wide range of lightness and chromaticity values. We measured the spectral reflectance factors of all samples by using a multi-angle spectrophotometer, the BYK mac-i multi-angle spectrophotometer, at the same six geometries that are also present in the Byko-spectra effect light booth.

As illustrated in Fig. 7, this set of samples includes RGB and yellow as well as achromatic samples. The samples are relatively small, 10 cm × 8.5 cm, thus minimizing the influence of the rather inhomogeneous illumination on the sample's platform of the light booth, as reported in Section 3. We used the spectral reflectance data to render the samples as they would appear inside the light booth. An example of such a rendering is shown in Fig. 8.

Sixteen observers (six females and 10 males aged between 20 and 57) participated in the experiment. All observers have normal color vision as confirmed by the Ishihara color vision test, and all have normal, or corrected to normal, visual acuity. We asked the observers to evaluate the color difference they perceived between the rendered samples and the real-world samples. We used two different psychophysical methods in order to detect experimental biases.

With the scoring method [22], we asked each observer to evaluate the perceived color differences between the rendered and the real-world samples by giving scores ranging from 0 to 5. A score of zero refers to the situation where the observer saw no color difference, or hardly any color difference. If the observer assessed the perceived color as large, meaning that the color of

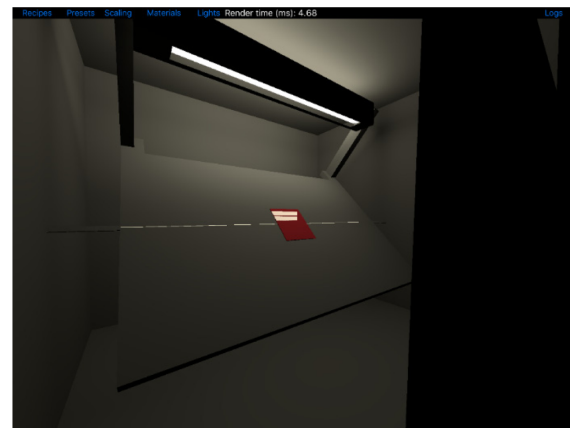


Fig. 8. Screenshot of the real-time rendering on an iPad of a flat high-gloss sample inside the virtual light booth.

Table 1. Descriptions of Scores for the Scoring Method

Score	Description
0	No/Hardly any difference
1	Small, negligible difference
2	Difference visible but still acceptable
3	Difference visible, doubtful match
4	Difference clearly visible, not correct match
5	Large difference, very bad match

the rendered sample was a very bad color match to the color of the real-world sample, then a score of five should be given. Similar descriptions for all intermediate score values were provided to the observers in the form of a table, which is reproduced here in Table 1.

The second psychophysical method that we used is the grayscale method [34]. In this case, we used a series of commercially available color chips known as the Society of Dyers and Colourists (SDC) grayscale [35]. These chips are widely used in color science as a reference set and serve to quantify the magnitude of color differences during visual tests. The grayscale consists of nine pairs of neutral gray-colored chips. The nine pairs are labeled 5, 4-5, 4, 3-4, 3, 2-3, 2, 1-2, and 1. Each pair has a lightness difference, the magnitude of which varies over the nine pairs. In the grayscale method, we asked the observer to decide which pair of grayscale chips represents a color difference that agrees best with the color difference observed between the rendered and the real-world samples.

For this test, we used chromatic paint samples, for which the color hardly varies with viewing and illumination directions. Therefore, we conducted the visual test for only one of the six geometries that can be selected in the Byko-spectra effect light booth. We selected the 45 as 45 geometry, in which light is incident from 45° with respect to the surface normal and where the observer watches the sample from a perpendicular direction. In future work for special effect coatings, we will use all six geometries of this light booth.

We conducted the visual experiment in a dark room to avoid any disturbance from light that was not included in the rendered lighting scene. We placed the real-world samples at the center

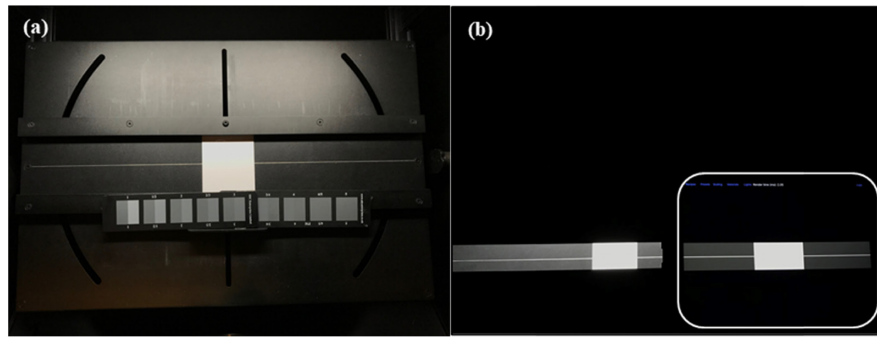


Fig. 9. Experimental setup for the visual test with (a) the grayscale located inside the light booth next to the real-world sample, (b) observer view of the real-world sample inside the light booth on the left, and rendered sample on the tablet display on the right. We added a white line to this photograph to outline the edge of the display.

of the sample platform inside the light booth. In case of the grayscale method, the color chip pairs were placed next to the real-world samples, as shown in Fig. 9(a). To show all nine gray pairs simultaneously, we needed two identical SDC grayscales because the full range of chips covers both the front and back sides of a single scale [36]. The tablet display was placed next to the viewing slit of the light booth, enabling the observer to directly compare the colors of the real-world and rendered samples with each other (by not placing the iPad inside the light booth, we avoid light reflected from internal light booth components to reach the display). This is illustrated in Fig. 9(b). In this setup, the viewer watches the display from a straight angle. This is important, because LCD displays as used in the iPad show a considerable variation in color on viewing angle, especially in luminance [37].

We covered the extreme parts of the viewing slit of the real-world light booth [far left and far right side in Fig. 9(b)] with a black fabric mask in order to match the dimensions of the virtual viewing slit, the dimensions of which are limited by the size of the tablet computer. Another mask was placed on the tablet to avoid stray light, which would be generated by the top and bottom edges of the virtual slit.

We subjected each observer to one training session. Before starting each session, we let observers adapt their color vision to the light booth for 2 min. Each observer assessed all 30 samples according to both psychophysical assessment methods and repeated the visual experiment three times in different sessions. Samples were presented in random order, with a different order being used at each session.

4. RESULTS

A. Intra-Observer Repeatability

We collected the visual scores from 16 observers on 30 samples using two different psychophysical methods and repeating each assessment independently three times. This gives a total of 2880 assessments.

We determined the intra-observer repeatability of the assessments by calculating the standard deviation of visual scores over the three repeated sessions for each sample and for each observer. When averaging the standard deviations over all samples and all observers (a simple average), we find that the intra-observer

repeatability is 0.59 for the scoring method and 0.43 for the grayscale method.

For the scoring method, Table 1 shows that the possible scores ranging from 0 to 5 with steps of 1 unit. The intra-observer repeatability of 0.59 units that we found is smaller than the quantification limit of the method and is therefore considered to indicate good intra-observer repeatability. For the grayscale method, the SDC grayscales vary from 1 to 5 with steps of 0.5 units. Since we found an intra-observer repeatability of 0.43 in this case, which is smaller than the quantification limit, we obtain good intra-observer repeatability with this method.

We note that the intra-observer repeatability for the two psychophysical methods is very similar to each other when calculated relative to the total range of attainable scores. For the scoring method, the intra-observer repeatability covers 0.59 from the total scale of five units, i.e., 11.8%. For the grayscale method, it covers 0.43 from a scale with a range of four units, i.e., 10.8%. Therefore, after normalization, both psychophysical methods give very similar results on intra-observer repeatability.

B. Inter-Observer Reproducibility

With the term reproducibility, we refer to the alignment between different observers. This was determined from the results of the visual assessments by first calculating for each sample the absolute difference between the average assessment for a particular observer and the average of all assessments from all observers. By taking the average over all samples, we obtain the inter-observer reproducibility, i.e., the average absolute deviation in the score of an observer with respect to the average score of all observers.

In this way, we found an inter-observer reproducibility of 0.74 units for the scoring method and 0.57 units for the grayscale method. It is not surprising that we find the inter-observer reproducibility to be larger than the intra-observer repeatability: in most psychophysical experiments, observers tend to agree better with their own earlier assessments than with assessments from other observers. Here, we found that inter-observer reproducibility is only slightly larger than the intra-observer repeatability, which indicates a low ambiguity between observers on the visual experiment.

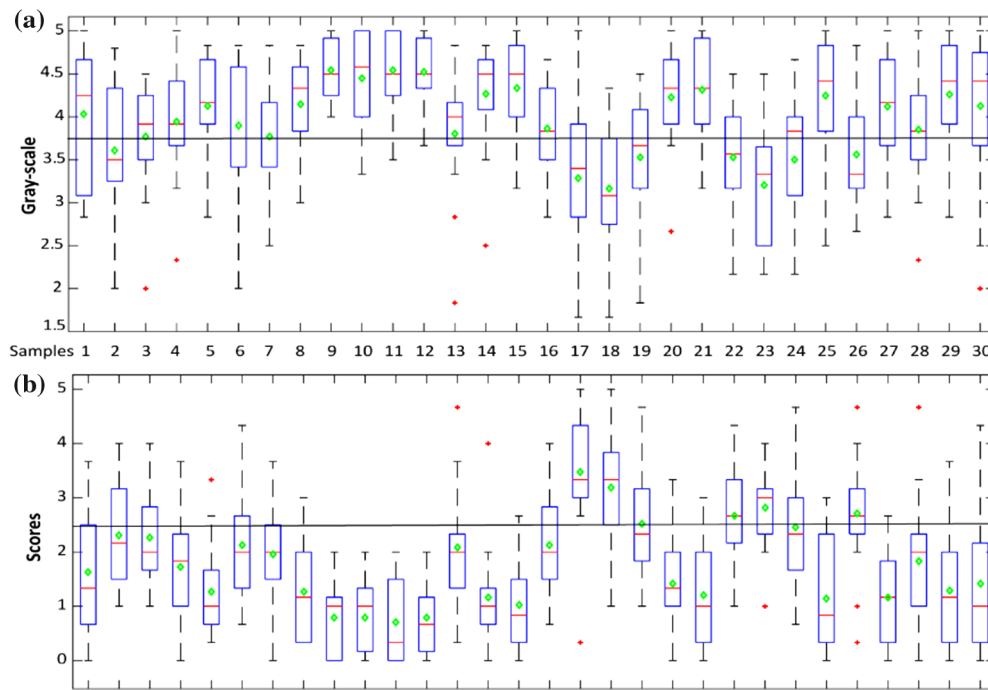


Fig. 10. Results of visual experiments. The red pluses indicate the outliers. Green pluses represent the mean value. Horizontal line illustrates the acceptance threshold for each method. Acceptable color match is obtained for grayscale values above the threshold value in the top graph and for scores below the threshold value in the bottom graph.

For the scoring method, the inter-observer reproducibility is smaller than the quantification limit of this method, whereas for the grayscale method it is only slightly larger. After normalization, the inter-observer reproducibility covers 14.8% of the scale for the scoring method and 14.3% for the grayscale method. We also conclude that for inter-observer reproducibility both psychophysical methods give very similar results.

C. Perceived Color Match

The collected visual scores provide quantitative information on the absolute color match of the rendering framework, as perceived by the observers. For the scoring method, the average visual score over all samples and all observers is 1.78. According to the descriptions listed in Table 1, this denotes a perceived color difference between “small, negligible difference” and “difference visible but still acceptable” as closer to the second description than to the first. This indicates that the color match is visually acceptable for the average sample. This color match is much better than what is found when using the default sRGB color encoding space. In an earlier publication, we used the same scoring method and the same definitions for visual scores as used here to evaluate the color reproduction of the sRGB method. For an iPad Air 2 display, we then found an average score of 3.6 to 4.6, i.e., assessed to lie between “difference visible; doubtful match” and “difference clearly visible; not correct match” [22].

For the sRGB method on an iPad Air 2, our previous results showed that the average score was smaller than three in only 34% of the cases or even less, depending on the ambient lighting [26]. These scores may be considered as referring to cases with a color reproduction accuracy that is reasonable or better. In

the current investigation, the rendering framework with the MDCIM model gives the same range of scores for 93% of the cases. Here, we chose a tighter threshold at 2.5 units to analyze our results, which is halfway between “difference visible but still acceptable” and “difference visible, doubtful match.” In 80% of the assessments, the color match was judged as being better than this threshold value.

Figure 10 shows the results of the visual tests with the scoring method in the form of a so-called box plot. Mean values are indicated by green pluses, red lines represent median values, and red pluses denote outliers in the data. The boxes extend between the first and third quartile values, and dashed lines connect the boxes to minimum and maximum values.

The second psychophysical method that we used in the visual tests utilizes the SDC grayscale. The labels of its chip pairs are labeled as 5, 4-5, 4, 3-4, 3, 2-3, 2, 1-2, and 1, with progressively increasing color differences between the pairs. For our numerical analysis, we relabeled them as the numerical values 5, 4.5, 4.0, ..., 1.0 in the same order.

With the grayscale method, we then find that the average color match of the rendering pipeline is judged to be 4.0 units. This is equivalent to a color difference of $\Delta E_{00} = 1.6$ in terms of CIEDE2000 units [36,38]. This corresponds to a small visible color difference, which agrees with the findings obtained with the other psychophysical test method. The range of visual scores obtained is shown as a box plot in Fig. 10 as well.

For the grayscale method, we chose a tolerance threshold of 3.75 units, as shown in Fig. 10. This threshold is equivalent to a color difference of $\Delta E_{00} = 2.0$ [36,38]. With this threshold value, we found that 77% of the samples have an acceptable color match. This percentage agrees well with the

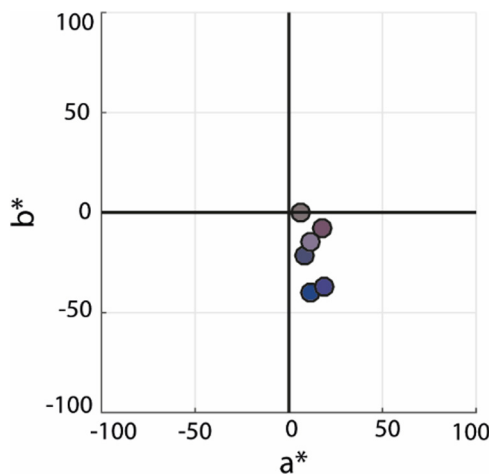


Fig. 11. Chromaticity diagrams with the color coordinates a^* , b^* of samples for which rendering was not acceptable according to the thresholds for the grayscale and scoring methods.

result obtained with the scoring method. This confirms that both psychophysical methods produce very similar results.

For the scoring method, six from the 30 samples (i.e., 20%) are not accepted by the threshold for that method. From the seven samples that are not accepted by the threshold in the grayscale method, six are the same as the unaccepted samples for the scoring method. The a^* , b^* color coordinates of the unaccepted samples from both methods are illustrated in Fig. 11. This graph shows that a relatively poor color match is obtained for samples with $a^* > 0$ and $b^* < 0$. A comparison of Fig. 11(b) with the color coordinates of all samples included in this investigation, as shown in Fig. 7(b), suggests that the color reproduction is more critical for reddish blue samples.

5. CONCLUSIONS

We described a recently developed rendering framework that should allow an improved color reproduction. The rendering pipeline makes use of data with 16 spectral bands, which were derived from spectrophotometric measurements from samples (spectral reflectance) to be rendered, light sources (spectral, angular, and spatial distribution of the radiant flux) and other secondarily relevant features of the scene that might affect appearance. The observation scene was normalized to be the perspective from the viewing slit of a Byko-spectra effect light booth, and it was simulated by the pipeline. We also took into account the colorimetric specifications of the rendering display (iPad5) by applying the recent device-specific MDCIM model. We evaluated the color reproduction of the new rendering framework by psychophysical tests using spectrophotometric measurements of 30 chromatic paint samples. In a visual test, 16 observers compared the color of each sample with the color of its rendered representation inside the virtual light booth on an iPad5 display. We collected a total of 2880 visual assessments, using two different psychophysical methods (scoring and grayscale). The intra-observer repeatability is smaller than the quantification limits of both methods, and the inter-observer repeatability is almost identical (14.8% and 14.3%). This shows that in only a few samples is there an inconsistency

between observers, or between repeated assessments by the same observer, about the color reproduction accuracy. Both methods show that the poorest color reproduction is obtained for samples with reddish blue colors. In 80% of the assessments, the color reproduction score is below our acceptance threshold. The current results are much better than what we found in a previous investigation when using the default sRGB color encoding space.

The results reported here for the rendering framework consider only solid color samples. In future research, we will investigate the inclusion in this framework of specific angular distributions of radiant flux and reflectance for effect coatings.

Funding. Ministerio de Economía y Competitividad (FPI BES-2016-077325, RTI2018-096000-B-I00).

Acknowledgment. The authors thank the Ministry of Economy and Competitiveness. The authors also thank the Ministry of Economy and Competitiveness for the pre-doctoral fellowship.

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. J. Dorsey and F. Sillion, *Digital Modeling of Material Appearance* (Morgan Kaufmann Elsevier, 2008).
2. Y. Dong, S. Lin, and B. Guo, *Material Appearance Modeling: A Data-Coherent Approach* (Springer, 2013).
3. M. Bordegoni and C. Rizzi, *Innovation in Product Design, From CAD to Virtual Prototyping* (Springer, 2011).
4. M. Haindl and J. Filip, *Accurate Material Appearance Measurement, Representation and Modeling*, Advances in Computer Vision and Pattern Recognition (Springer, 2013).
5. R. Martín, M. Weinmann, and M. Hullin, "Digital transmission of subjective material appearance," *J. WSCG* **25**, 57–66 (2017).
6. A. Ferrero, J. Campos, E. Perales, A. Rabal, F. Martínez-Verdú, A. Pons, E. Chorro, and M. L. Hernanz, "Measuring and specifying goniochromatic colors," in *23rd Congress of the International Commission for Optics*, Santiago de Compostela, Spain (2014).
7. F. Maile, G. Pfaff, and P. Reynders, "Effect pigments—past, present and future," *Progr. Org. Coat.* **54**, 150–163 (2005).
8. E. Kirchner, I. van der Lans, A. Ferrero, J. Campos, F. M. Martínez-Verdú, and E. Perales, "Fast and accurate 3D rendering of automotive coatings," in *23rd Color and Imaging Conference* (Society for Imaging Science and Technology, 2015), pp. 154–160.
9. T. Golla and R. Klein, "Interactive interpolation of metallic effect car paints," in *Vision, Modeling and Visualization, Eurographics* (2018).
10. E. Kirchner, "Texture measurement, modeling, and computer graphics," in *Encyclopedia of Color Science and Technology*, M. R. Luo, ed. (Springer, 2015).
11. J. Günther, T. Chen, M. Goesele, I. Wald, and H.-P. Seidel, "Efficient acquisition and realistic rendering of car paint," in *Proceedings of Vision, Modeling and Visualization*, Erlangen, Germany (2005), pp. 487–494.
12. M. Rump, G. Müller, R. Sarlette, D. Koch, and R. Klein, "Photo-realistic rendering of metallic car paint from image-based measurements," in *Eurographics/Computer Graphics Forum*, Crete, Greece (2008), 27, pp. 527–536.
13. C. Shimizu and G. Meyer, "A computer aided color appearance design system for metallic car paint," *J. Imaging Sci. Technol.* **59**, 304031 (2015).
14. A. I. Ruppertsberg and M. Bloj, "Rendering complex scenes for psychophysics using RADIANCE: how accurate can you get?" *J. Opt. Soc. Am. A* **23**, 759–768 (2006).
15. J. A. Ferwerda, S. Westin, R. Smith, and R. Pawlicki, "Effects of rendering on shape perception in automobile design," in *ACM Symposium on Applied Perception in Graphics and Visualization* (2004), pp. 107–114.

16. M. Nimier-David, D. Vicini, T. Zeltner, and W. Jakob, "Mitsuba 2: a retargetable forward and inverse renderer," *ACM Trans. Graph.* **38**, 1–17 (2019).
17. J. A. Ferwerda, "ImpastoR: a realistic surface display system," *Vis. Res.* **109**, 166–177 (2015).
18. E. Valenza, *Blender Cycles: Materials and Textures Cookbook* (Packt, 2015).
19. C. Nguyen, M.-H. Kyung, J.-H. Lee, and S.-W. Nam, "A PCA decomposition for real-time BRDF editing and relighting with global illumination," *Comput. Graph. Forum* **29**, 1469–1478 (2010).
20. A. Ferrero, E. Perales, A. Rabal, J. Campos, F. Martínez-Verdú, and A. Pons, "Color representation and interpretation of special effect coatings," *J. Opt. Soc. Am. A* **31**, 436–447 (2014).
21. A. Ferrero, J. Campos, E. Perales, F. Martínez-Verdú, I. van der Lans, and E. Kirchner, "Global color estimation of special-effect coatings from measurements by commercially available portable multiangle spectrophotometers," *J. Opt. Soc. Am. A* **32**, 1–11 (2015).
22. E. Kirchner, I. van der Lans, F. Martínez-Verdú, and E. Perales, "Improving color reproduction accuracy of a mobile liquid crystal display," *J. Opt. Soc. Am. A* **34**, 101–110 (2017).
23. X. Gao, E. Khodamoradi, L. Guo, X. Yang, S. Tang, W. Guo, and Y. Wang, "Evaluation of colour appearances on smartphones," in *Midterm Meeting of the International Colour Association*, Tokyo, Japan (AIC, 2015).
24. E. Kirchner, I. van der Lans, P. Koeckhoven, K. Huraibat, F. M. Martínez-Verdú, E. Perales, A. Ferrero, and J. Campos, "Real-time accurate rendering of color and texture of car coatings," in *IS&T International Symposium on Electronic Imaging* (2019).
25. B. A. Darling, J. A. Ferwerda, R. S. Berns, and T. Chen, "Real-time multi-spectral rendering with complex illumination," in *IS&T/SID 19th Color Imaging Conference* (2011), pp. 345–351.
26. "Tolerances for automotive paints. Part 2: Goniochromatic paints," DIN 6175-2 (2001), pp. 1–8.
27. F. M. Martínez-Verdú, E. Perales, V. Viqueira, E. Chorro, F. Burgos-Fernández, and J. Pujol, "Comparison of colorimetric features of some current lighting booths for obtaining a right visual and instrumental correlation for gonio-apparent coatings and plastics," in *CIE, Lighting Quality and Energy Efficiency* (2012), pp. 692–705.
28. K. Huraibat, E. Perales, V. Viqueira, E. Kirchner, I. van der Lans, A. Ferrero, and J. Campos, "Byko-spectra effect light booth simulation for digital rendering tool," in *XII Congreso Nacional del Color*, Linares, Spain (2019), pp. 45–48.
29. K. Huraibat, E. Perales, F. M. Verdú, E. Kirchner, I. van der Lans, A. Ferrero, and J. Campos, "Characterization of Byko-spectra light booth for digital simulation in a rendering tool," in *29th Quadrennial Session of the CIE*, Washington DC, USA (2019).
30. Blender Foundation, 2002, <https://www.blender.org>.
31. "Standard test method for multiangle color measurement of metal flake pigmented materials," ASTM. E 2194-14 (ASTM, 2014).
32. "Standard test method for multiangle color measurement of interference pigments," ASTM. E 2539-14 (ASTM, 2014).
33. E. Heitz, J. Dupuy, S. Hill, and D. Neubelt, "Real-time polygonal-light shading with linearly transformed cosines," *ACM Trans. Graph.* **35**, 1–8 (2016).
34. E. Montag and D. Wilber, "A comparison of constant stimuli and gray-scale methods of color difference scaling," *Color Res. Appl.* **28**, 36–44 (2003).
35. "SDC grey scale," <https://www.sdcenterprises.co.uk/products/sdc-assessment-aids/grey-scale/>.
36. SDC, *Methods of Tests for Colour Fastness of Textiles and Leather*, 5th ed. (The Society of Dyes and Colorists, 1990).
37. W. Cummings and T. Fiske, "Simplified ambient performance assessment for mobile displays using easy measurements," *SID Digest*. **45**, 528–531 (2014).
38. "Standard test method for pilling resistance and other related surface changes of textile fabrics: random tumble pilling tester method," ASTM D 3512 (2016).